NIR ITALIA 2022
9th National Symposium of NIR Spectroscopy
Izola, Slovenia, 7-9 June 2020

Strategies for non-linear modelling of NIR data

Ruggero Guerrini[1], Alessandra Biancolillo[2], Federico Marini[1*]

[1] Dept. Chemistry, University of Rome La Sapienza, Rome, Italy
[2] Dept. of Physical and Chemical Sciences, University of L'Aquila, Coppito, Italy
*Corresponding author (federico.marini@uniroma1.it)

The growth of modern high-throughput instrumentation, which makes avaialble the relatively rapid characterization of samples by often multiplatform and/or multidimensional approaches, results in data that, if not always "big", anyway have a high extent of granularity, with different degrees of similarity and a general inhomogeneity due to many sources of variation, other than those of interest. As a consequence, linear approaches often result inadequate both for regression and classification problems. A possible and obvious solution is to resort to non-linear modelling strategies, possibly where the degree of non-linearity can be tuned depending on the specific characteristics of the data set and kernel-based models, such as SVM or kernel-PLS(-DA) (Rosipal, 2003) are often used for such purpose. On the other hand, local approaches (Centner and Massart, 1998), which build opportunistic models based only of a subset of instances being most similar to the sample(s) to be predicted, represent a valid tool to achieve high accuracy and to take into account the stratification of the data set. However, traditional approaches to local modeling, although having been already used successfully in various domains, have some drawbacks, such a not always clear validation strategy, the lack of specific interpretation tools, their suboptimal performances in the presence of highly irregular distributions and their being almost always limited to single data matrices.

In the present communication, different strategies for dealing with non-linearities or different degrees of heterogeneities for modelling spectroscopic data will be discussed. In particular, new approaches to local modeling will be presented, which try to address some of the limitations mentioned above, and that can be generalized to the cases when the collected data have a multi-way or multi-block structures.

Keywords: non-linear modeling, local regression, kernel partial least squares regression (KPLS), kernel partial least squares discriminant analysis (KPLS-DA), dissimilarity PLS

REFERENCES

Centner, V, Massart D.L., 1998. Optimization in locally weighted regression. Anal. Chem. 70, 4206-11.

Rosipal, R. 2003. Kernel Partial Least Squares for Nonlinear Regression and Discrimination. Neural Network World, 13, 291-300.

Strategies for il modellazione non-lineare di dati NIR

Ruggero Guerrini[1], **Alessandra Biancolillo**[2], **Federico Marini**[1*]

[1] Dept. Chemistry, University of Rome La Sapienza, Rome, Italy
[2] Dept. of Physical and Chemical Sciences, University of L'Aquila, Coppito, Italy
*Corresponding author (federico.marini@uniroma1.it)

Lo sviluppo della moderna strumentazione high-throughput, che rende possibile la caratterizzazione relativamente rapida dei campioni mediante approcci spesso multipiattaforma e/o multidimensionali, si traduce in dati che, se non sempre "big", hanno comunque un'elevata granularità, con gradi differenti di somiglianza e una generale disomogeneità dovuta a molte fonti di variazione, oltre a quelle di interesse. Di conseguenza, gli approcci lineari risultano spesso inadeguati sia per problemi di regressione che di classificazione. Una possibile e ovvia soluzione è ricorrere a strategie di modellazione non lineare, possibilmente in cui il grado di non linearità possa essere regolato a seconda delle caratteristiche specifiche del set di dati e dei modelli basati sul kernel, come SVM o kernel-PLS(- DA) (Rosipal, 2003) sono spesso utilizzati per tale scopo. D'altra parte, gli approcci locali (Centner e Massart, 1998), che costruiscono modelli opportunistici basati solo su un sottoinsieme di istanze più simili al/ai campione/i da prevedere, rappresentano un valido strumento per ottenere un'elevata accuratezza e per tenere conto della stratificazione del set di dati. Tuttavia, gli approcci tradizionali alla modellazione locale, pur essendo già stati utilizzati con successo in vari domini, presentano alcuni inconvenienti, come una strategia di validazione non sempre chiara, la mancanza di strumenti interpretativi specifici, le loro prestazioni subottimali in presenza di distribuzioni altamente irregolari e il loro essere quasi sempre limitato a singole matrici di dati.

Nella presente comunicazione verranno discusse diverse strategie per trattare le non linearità o diversi gradi di eterogeneità per la modellizzazione dei dati spettroscopici. In particolare verranno presentati nuovi approcci alla modellazione locale, che cercano di affrontare alcune delle limitazioni sopra menzionate, e che possono essere generalizzati ai casi in cui i dati raccolti hanno una struttura multi-way o multi-block.

Keywords: modellazione non-lineare, regression locale, kernel partial least squares regression (KPLS), kernel partial least squares discriminant analysis (KPLS-DA), dissimilarity PLS

REFERENCES

Centner, V, Massart D.L., 1998. Optimization in locally weighted regression. Anal. Chem. 70, 4206-11.

Rosipal, R. 2003. Kernel Partial Least Squares for Nonlinear Regression and Discrimination. Neural Network World, 13, 291-300.